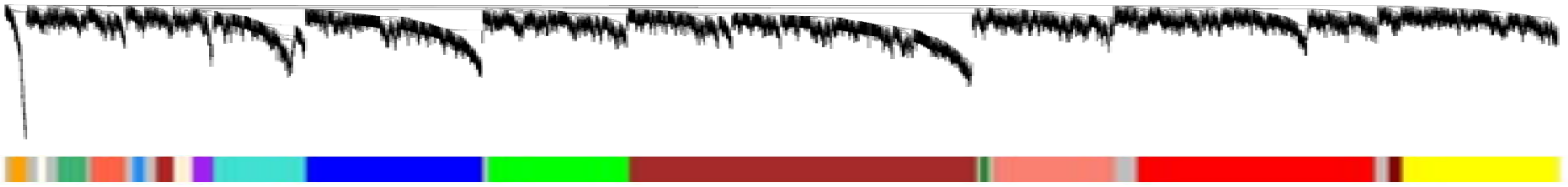
Functional Organization of the Transcriptome in Human Brain

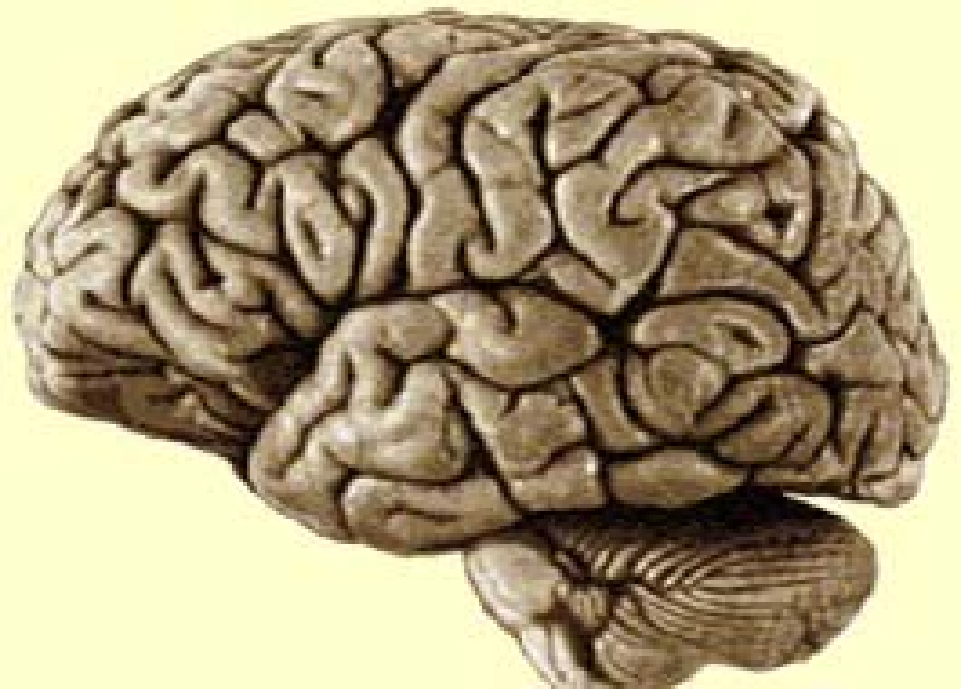
Michael C. Oldham

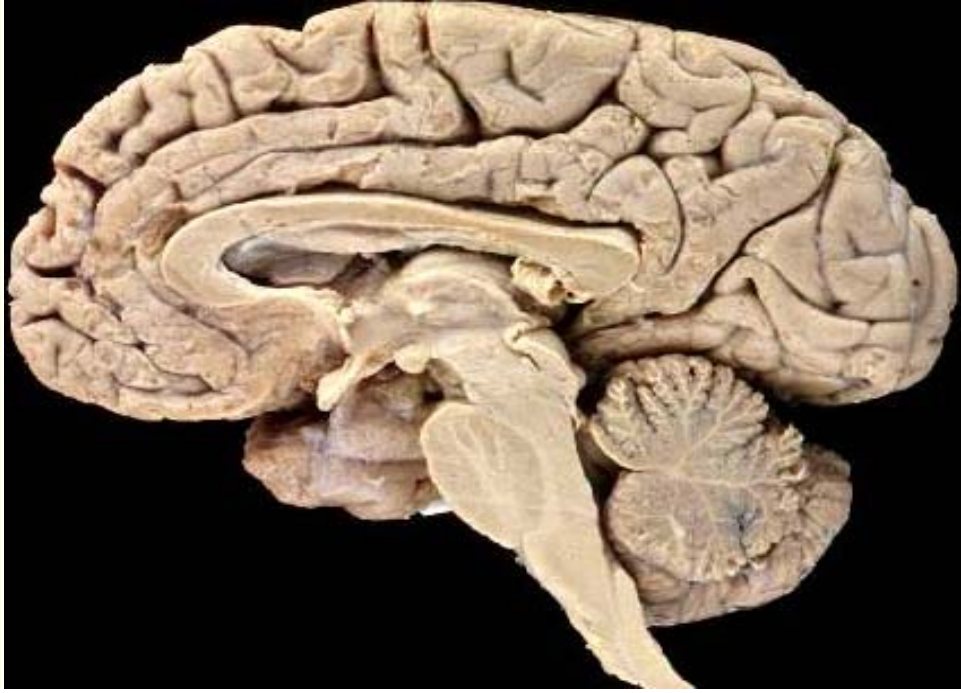
Laboratory of Daniel H. Geschwind, UCLA

BIOCOMP ‘08, Las Vegas, NV

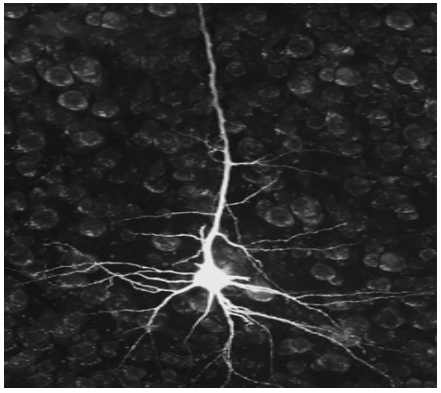
July 15, 2008



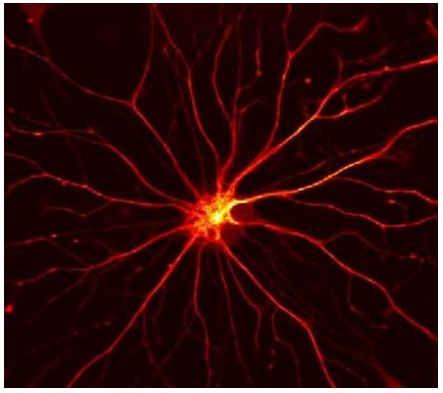




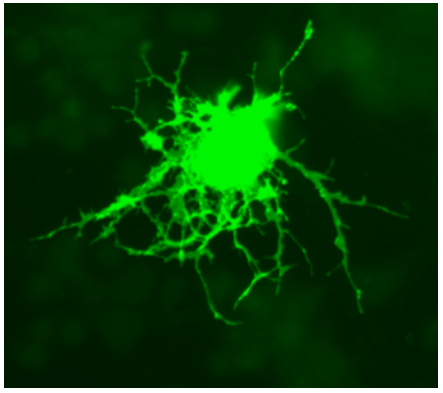
Neurons



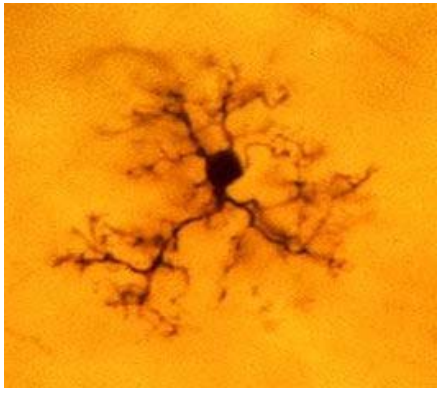
Astrocytes



Oligodendrocytes



Microglia



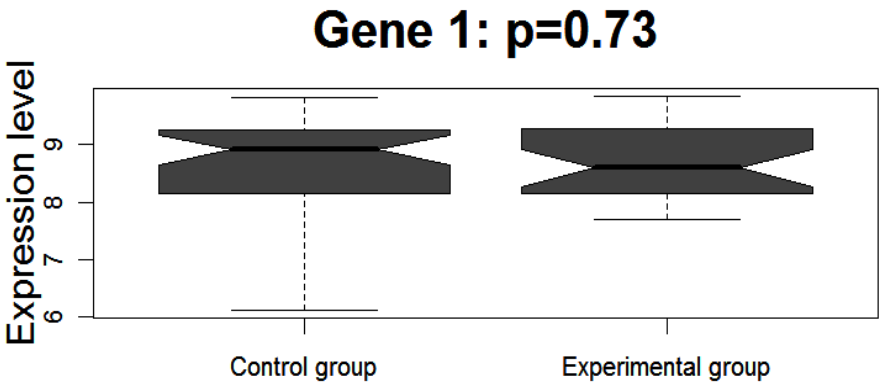
$\sim 10^{11}$

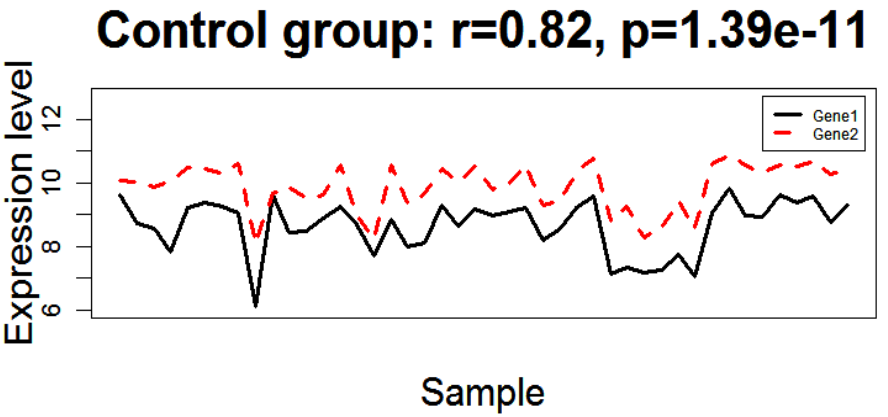
$\sim 10^{12}$

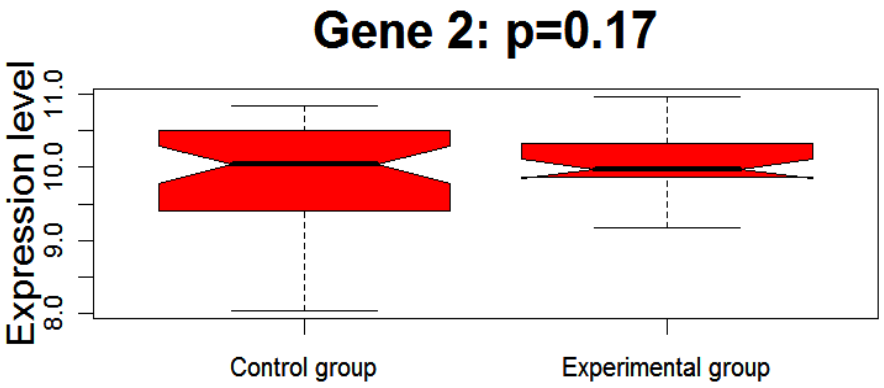
Microarray studies of the human brain face a number of challenges

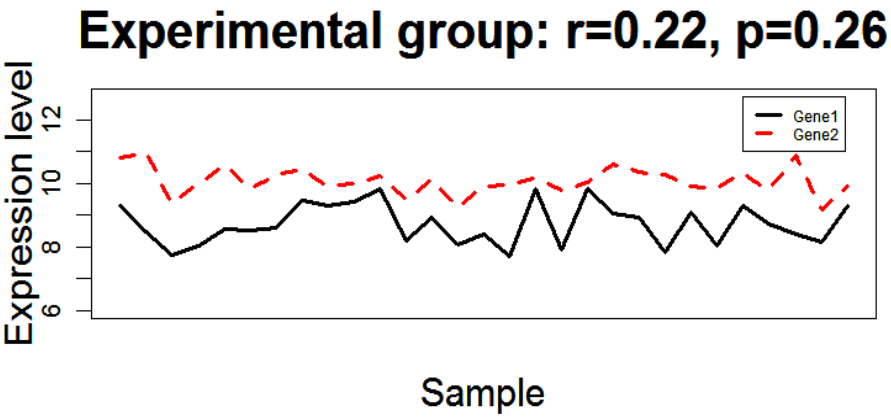
- Cellular heterogeneity
  - mRNA extracted from tissue homogenates
- Sample quality
  - Post-mortem tissue
- Sample size
  - Typically, <10 individuals per group
- Focus on differential expression

What's wrong with differential expression?



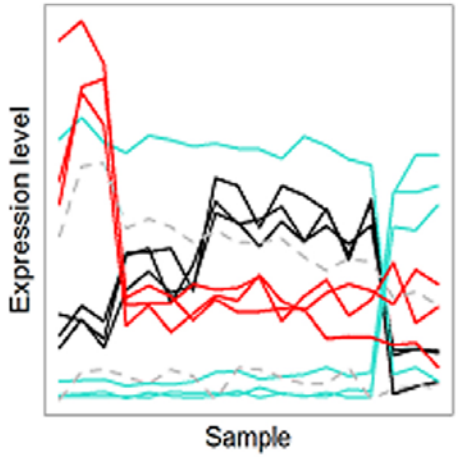






Methodological overview: WGCNA

A



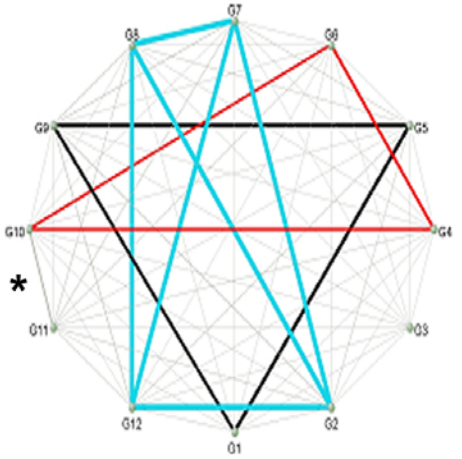
B

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
G1	1	-0.59	0.01	-0.47	0.96	-0.45	-0.56	0.55	0.94	-0.35	-0.01	-0.60
G2	-0.59	1	-0.28	-0.22	-0.67	-0.11	1.00	-0.99	-0.74	-0.38	-0.68	0.99
G3	0.01	-0.28	1	0.01	0.11	0.09	-0.30	0.29	0.07	0.23	0.40	-0.27
G4	-0.47	-0.22	0.01	1	-0.42	0.93	-0.22	0.27	-0.35	0.90	0.65	-0.20
G5	0.96	-0.67	0.11	-0.42	1	-0.40	-0.65	0.63	0.94	-0.26	0.13	-0.68
G6	-0.45	-0.11	0.09	0.93	-0.40	1	-0.11	0.16	-0.39	0.90	0.63	-0.08
G7	-0.56	1.00	-0.30	-0.22	-0.65	-0.11	1	-0.99	-0.73	-0.38	-0.69	0.99
G8	0.55	-0.99	0.29	0.27	0.63	0.16	-0.99	1	0.71	0.43	0.72	-0.99
G9	0.94	-0.74	0.07	-0.35	0.94	-0.39	-0.73	0.71	1	-0.20	0.17	-0.75
G10	-0.35	-0.38	0.23	0.90	-0.26	0.90	-0.38	0.43	-0.20	1	0.88	-0.35
G11	-0.01	-0.68	0.40	0.65	0.13	0.63	-0.69	0.72	0.17	0.88	1	-0.66
G12	-0.60	0.99	-0.27	-0.20	-0.68	-0.08	0.99	-0.99	-0.75	-0.35	-0.66	1

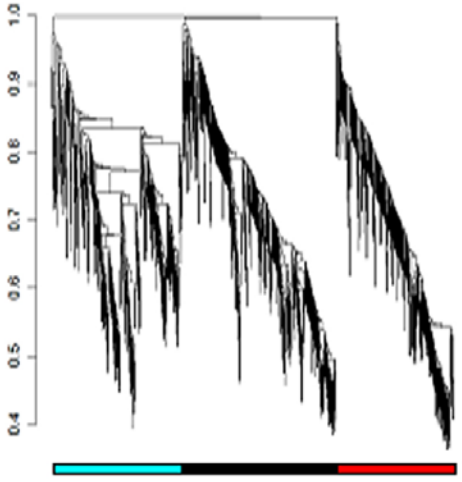
ß → Σ

Σ
k1
k2
k3
k4
k5
k6
k7
k8
k9
k10
k11
k12

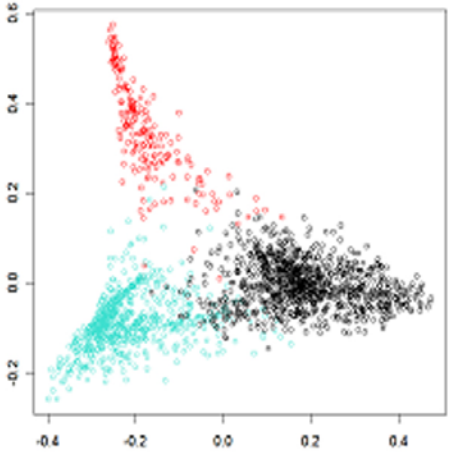
C



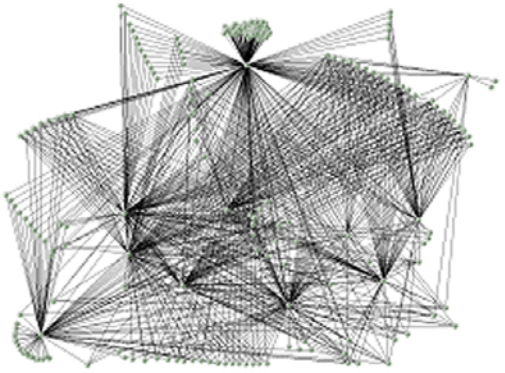
D



E



F



The data

Network 1 (U133 cortex): 67 samples, 67 individuals

Brain region	Reference	# samples	# outliers	# samples post QC
BA46	Iwamoto et al.*	34	5	20**
BA9, BA11	Ryan et al.	42 (31 BA9, 11 BA11)	8 (5 BA9, 3 BA11)	21 (13 BA9, 8 BA11)**
BA4, BA9	Hodges et al.	28 (16 BA4, 12 BA9)	2 (2 BA9)	26 (16 BA4, 10 BA9)

Network 2 (U95 cortex): 42 samples, 32 individuals

Brain region	Reference	# samples	# outliers	# samples post QC
BrA (x2), ACC, PrV, Prf, BrRH, Prm	Khaitovich et al.	21	7	12 (2 each region)
FP, aIT, IPL, MFG	Caceres et al.	7	5	2 (2 FP)
BA9	Enard et al.	6	2	2 (2 BA9)
BA9	Khaitovich et al.	3	2	1 (BA9)
BA10	Lu et al.	30	18	12 (BA10)
BA10	Iwamoto et al.*	15	2	13 (BA10)

Network 3 (U133 caudate nucleus): 27 samples, 27 individuals

Brain region	Reference	# samples	# outliers	# samples post QC
Caudate nucleus	Hodges et al.	32	5	27

Network 4 (U133 cerebellum): 24 samples, 24 individuals

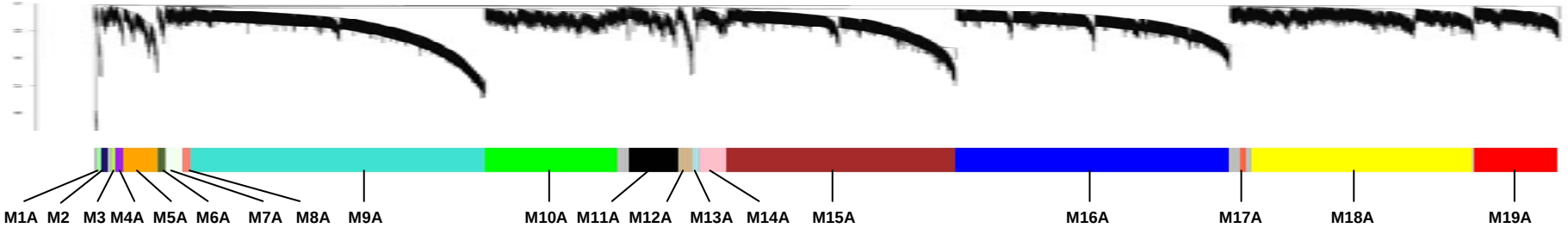
Brain region	Reference	# samples	# outliers	# samples post QC
Cerebellum	Hodges et al.	27	3	24

* Raw data obtained in collaboration with Dr. Kazuya Iwamoto and Dr. Tadafumi Kato at RIKEN.

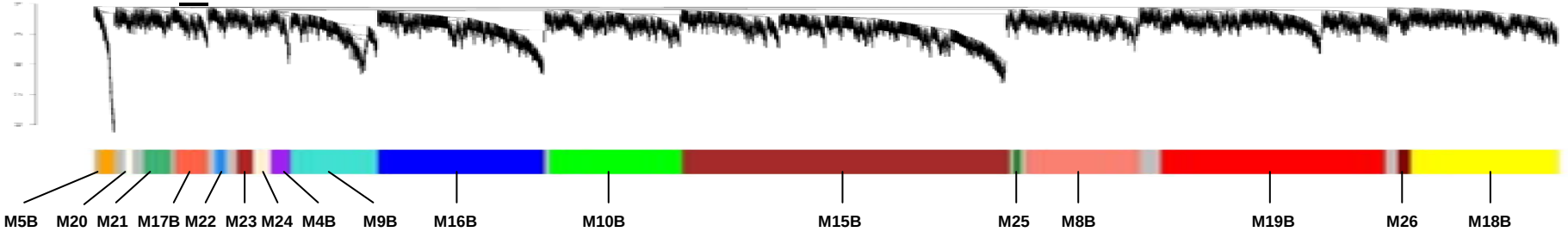
** Following outlier removal, there were 11 pairs of samples from the same individuals in Iwamoto et al. and Ryan et al. One unique sample per individual was retained.

The networks

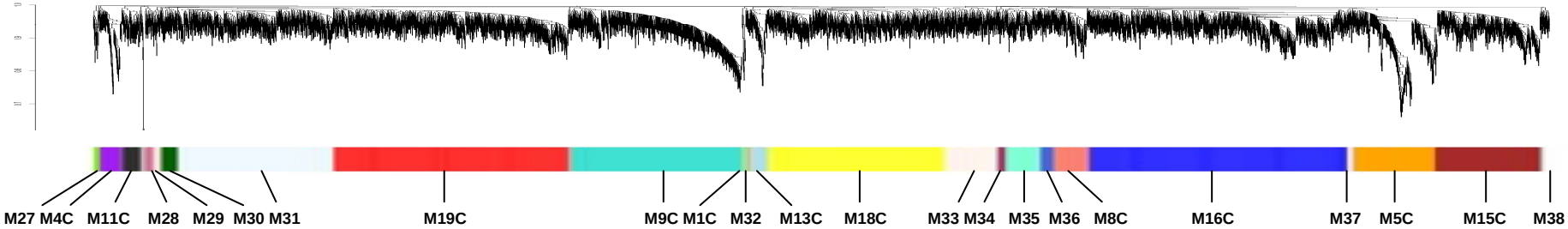
A: CTX



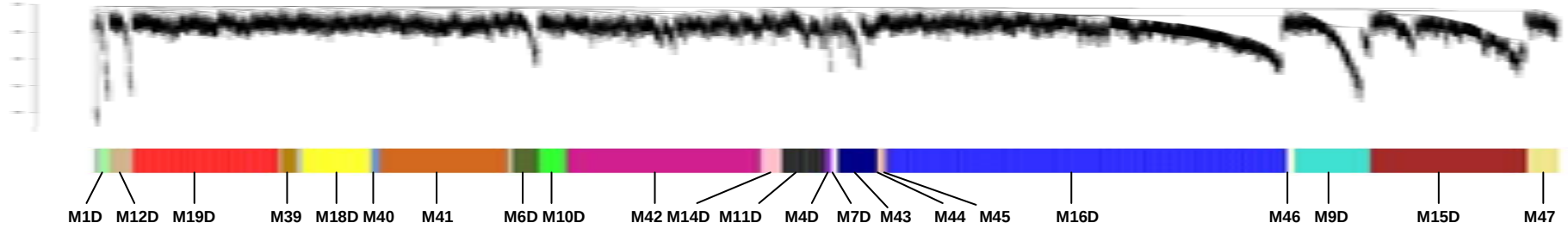
B: CTX_95



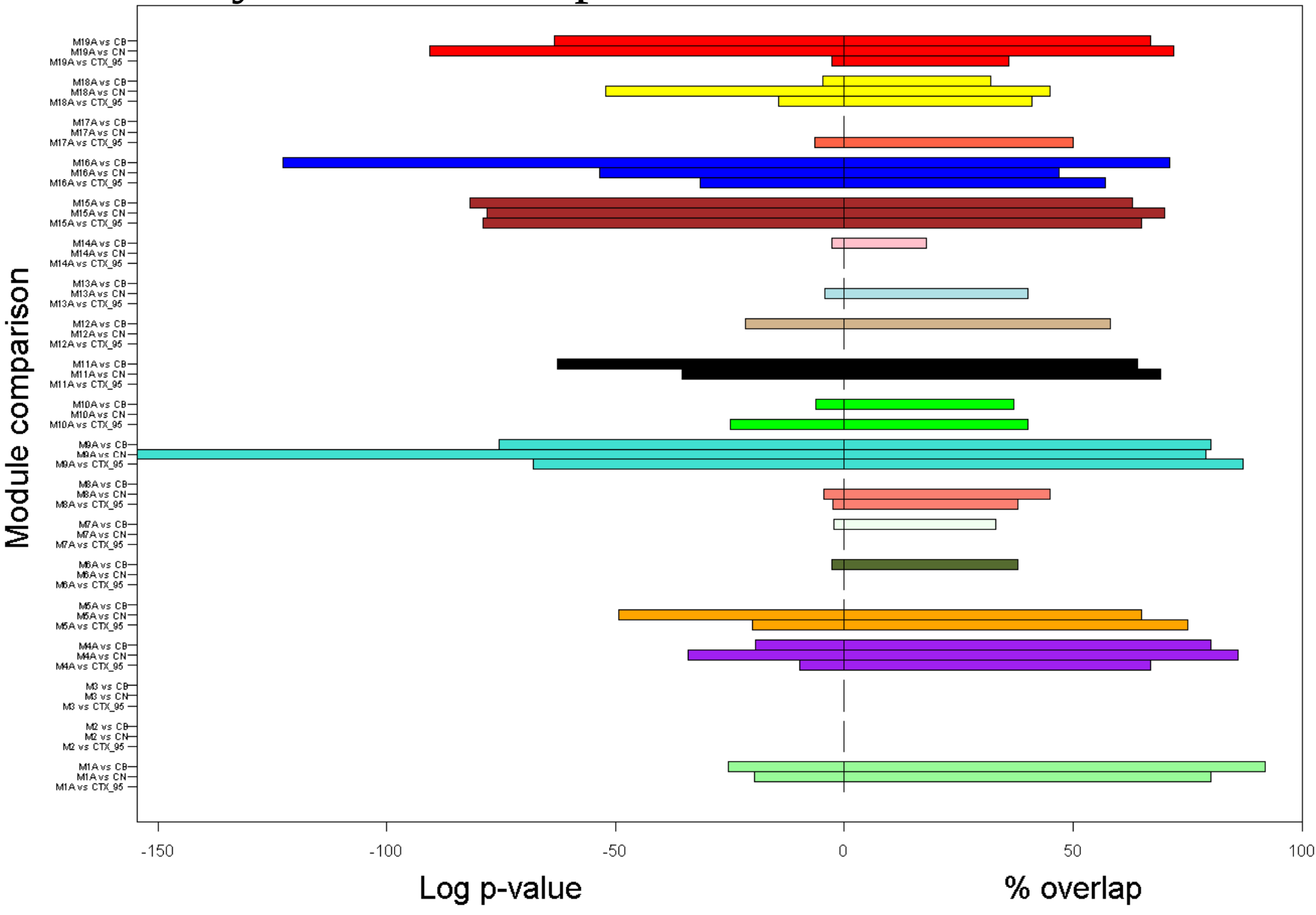
C: CN



D: CB

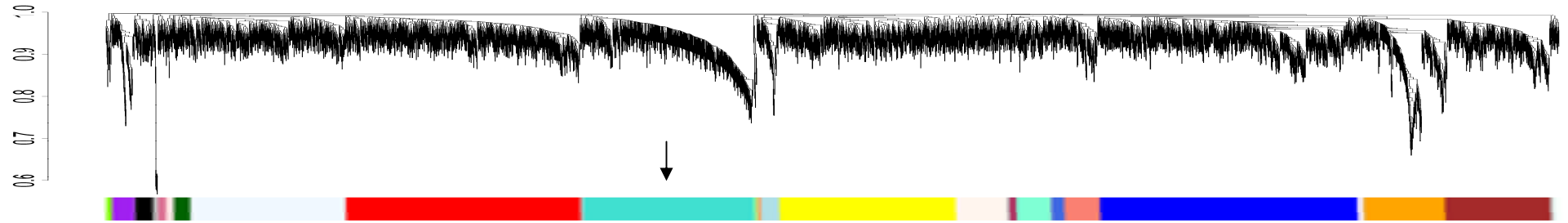


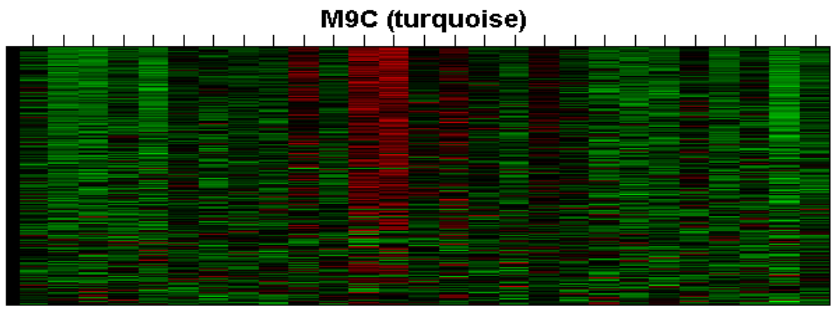
Many modules are preserved across networks

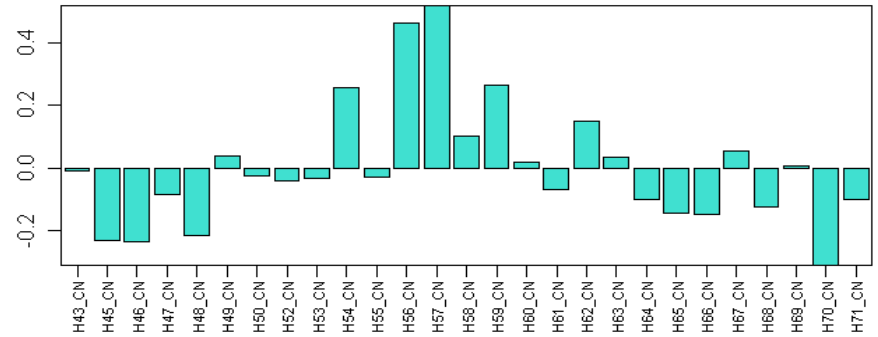


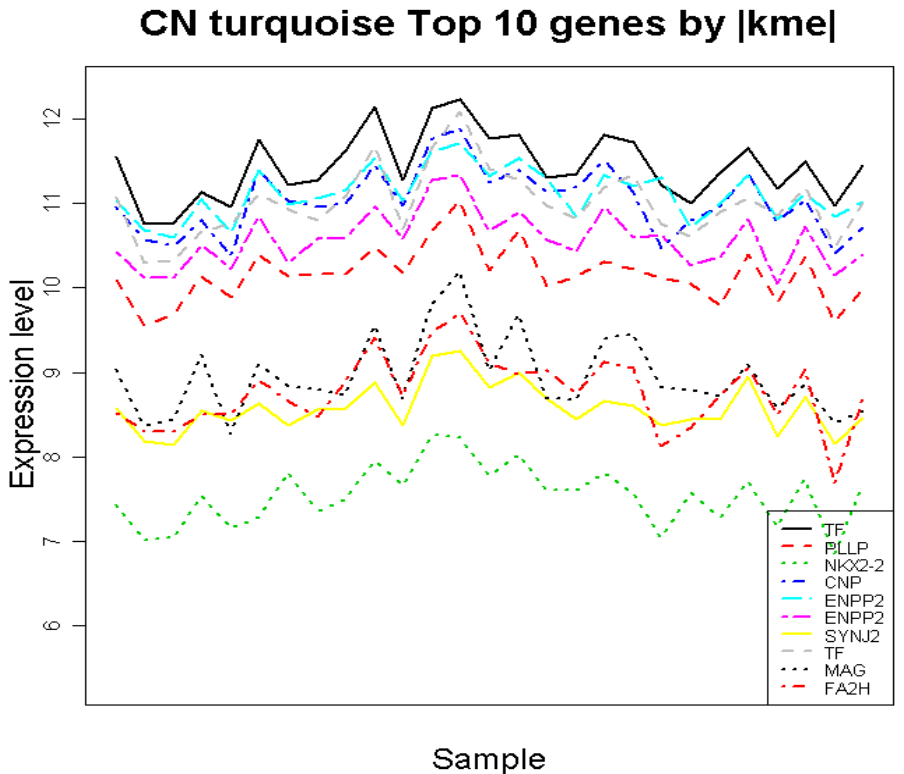
Quantifying module membership with k_{me}

- k_{me} is the Pearson correlation between the expression level of a given probe set and a given eigengene, e.g.:

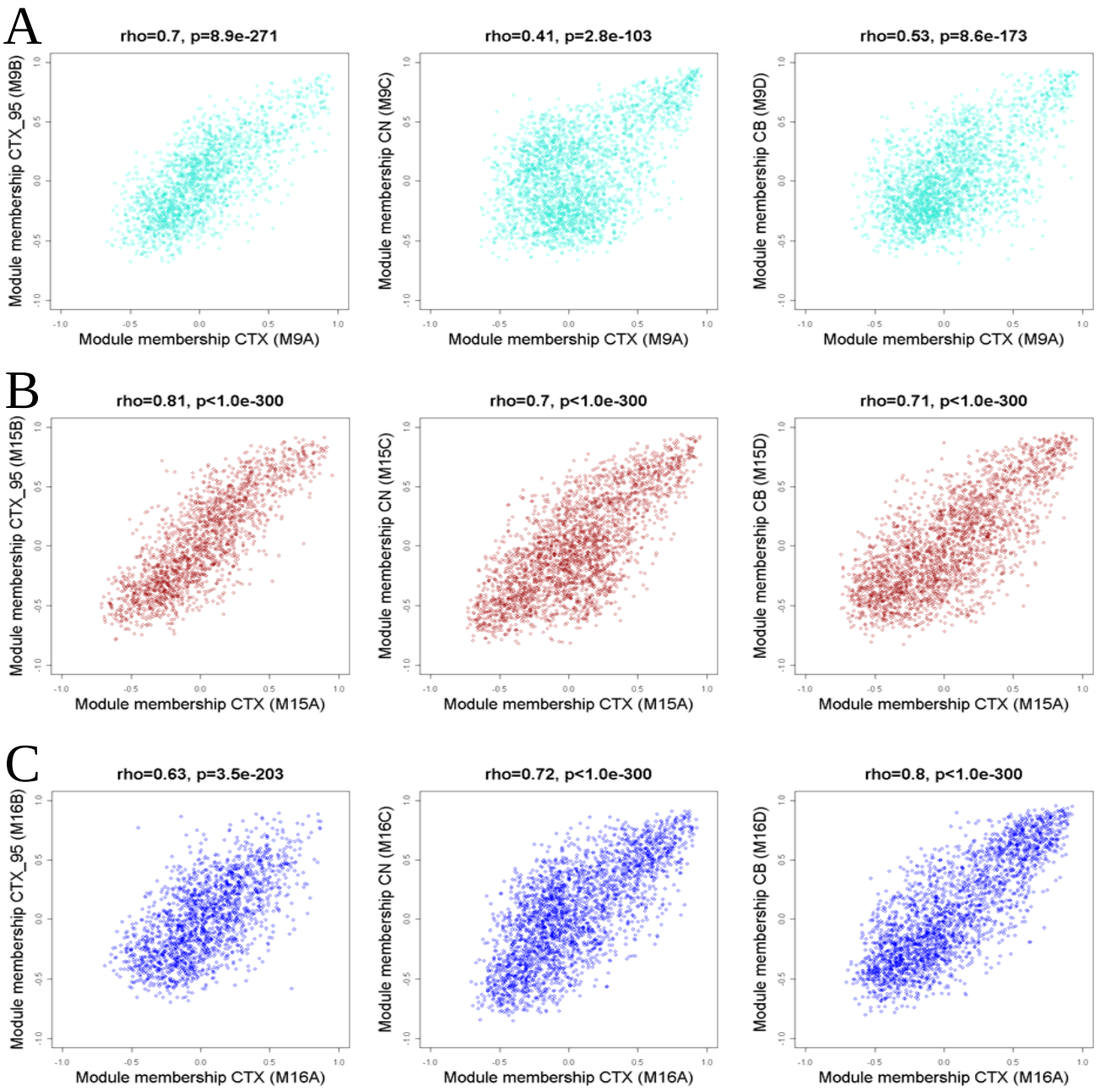




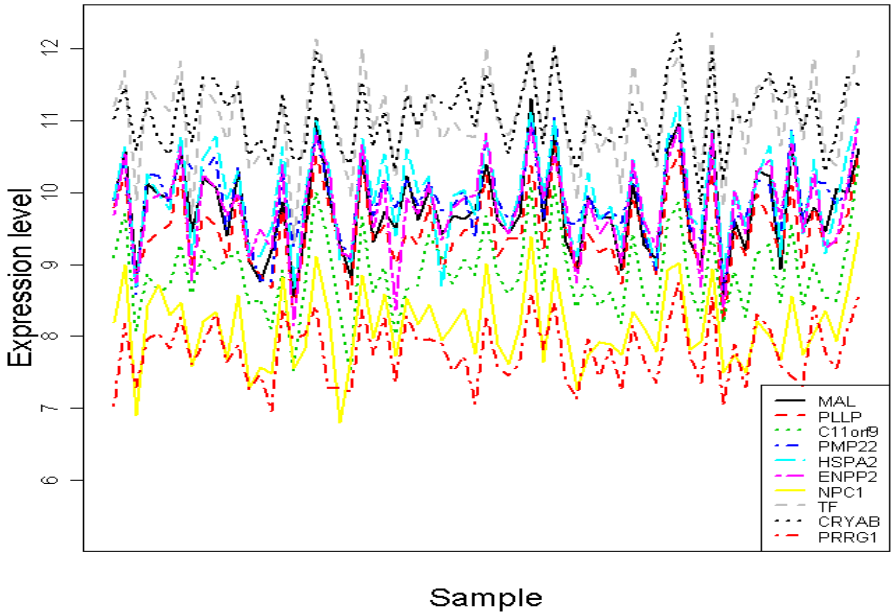




k_{me} is significantly correlated across networks

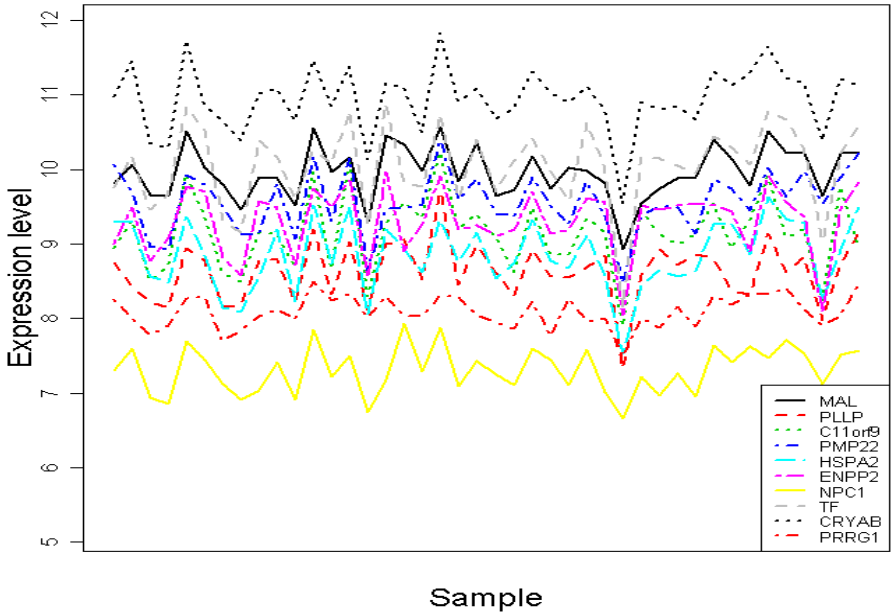


Cortex



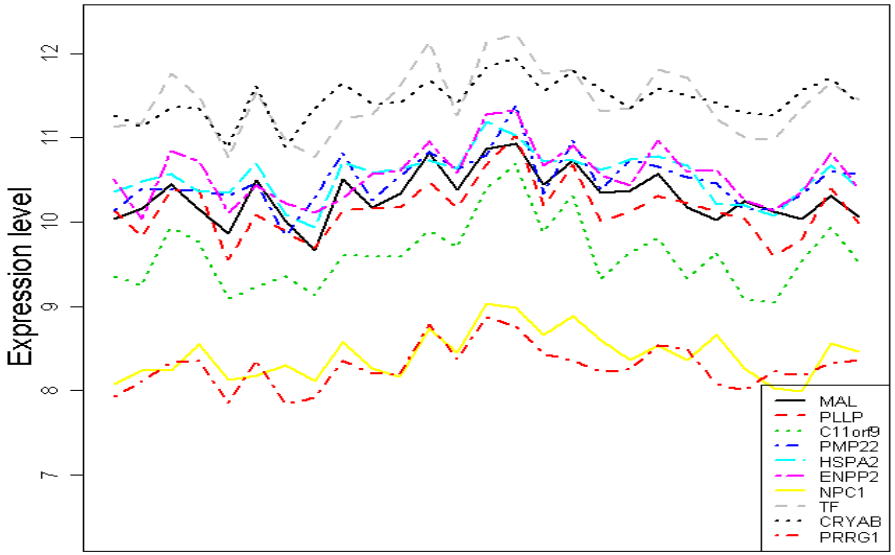
Sample

Cortex_95



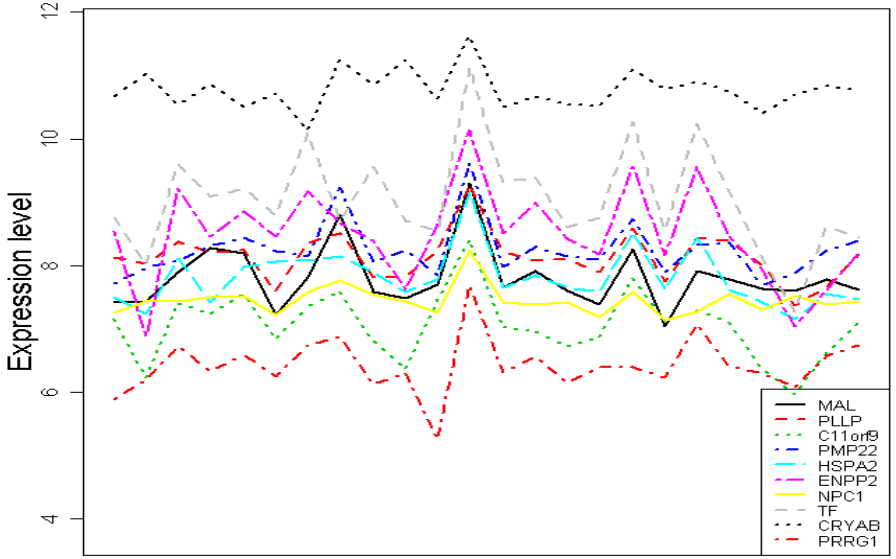
Sample

Caudate nucleus



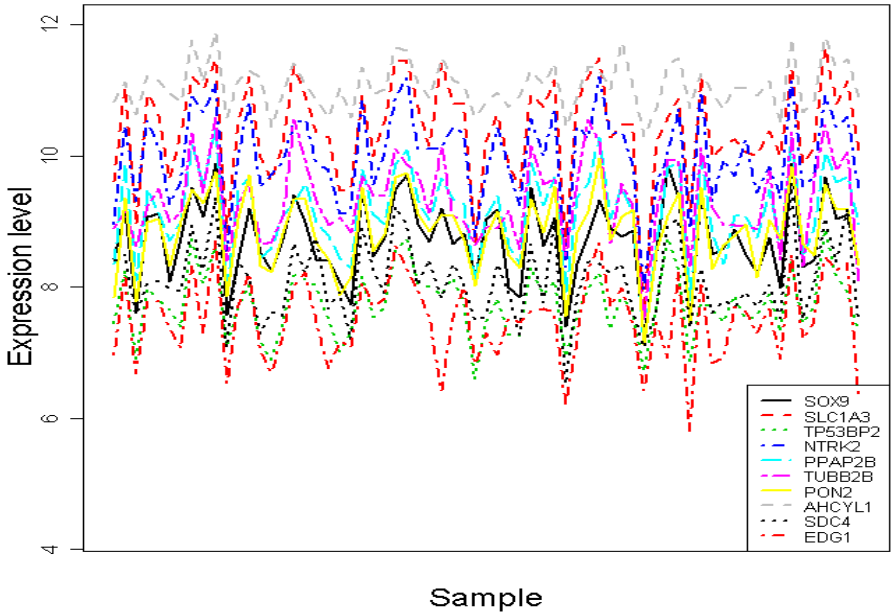
Sample

Cerebellum

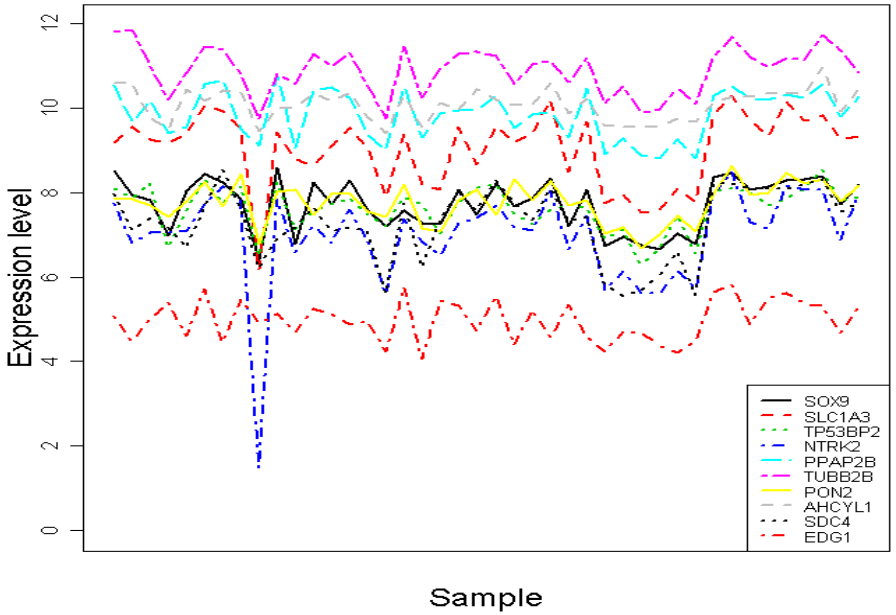


Sample

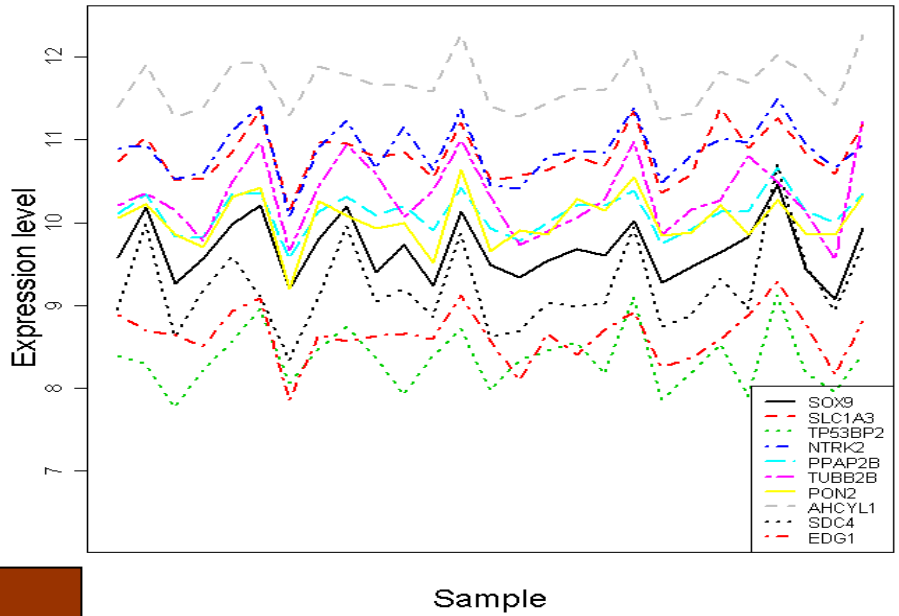
Cortex



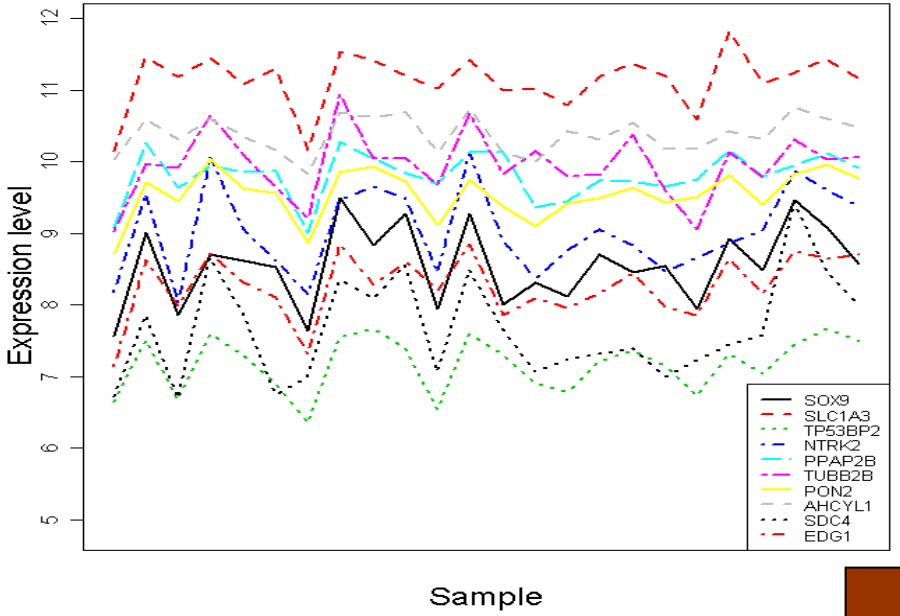
Cortex_95



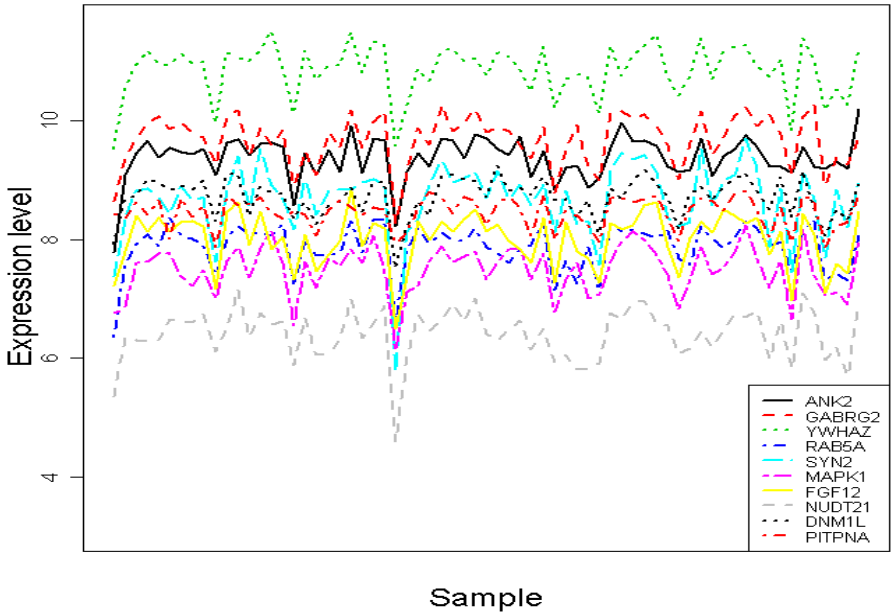
Caudate nucleus



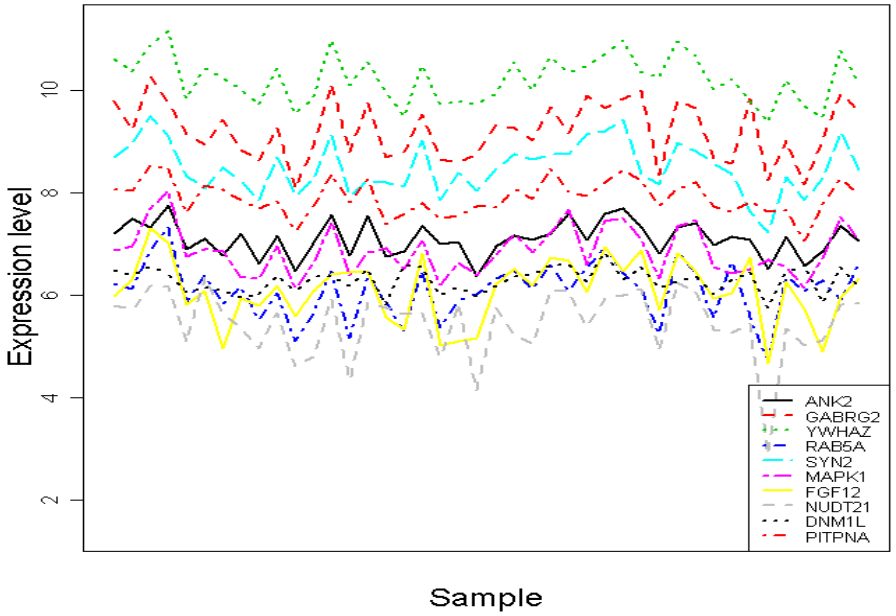
Cerebellum



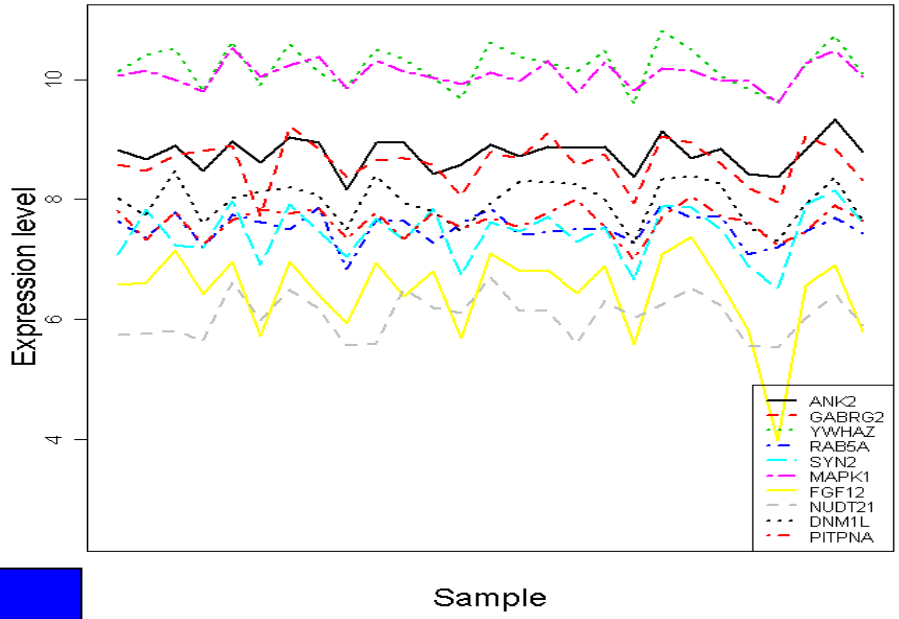
Cortex



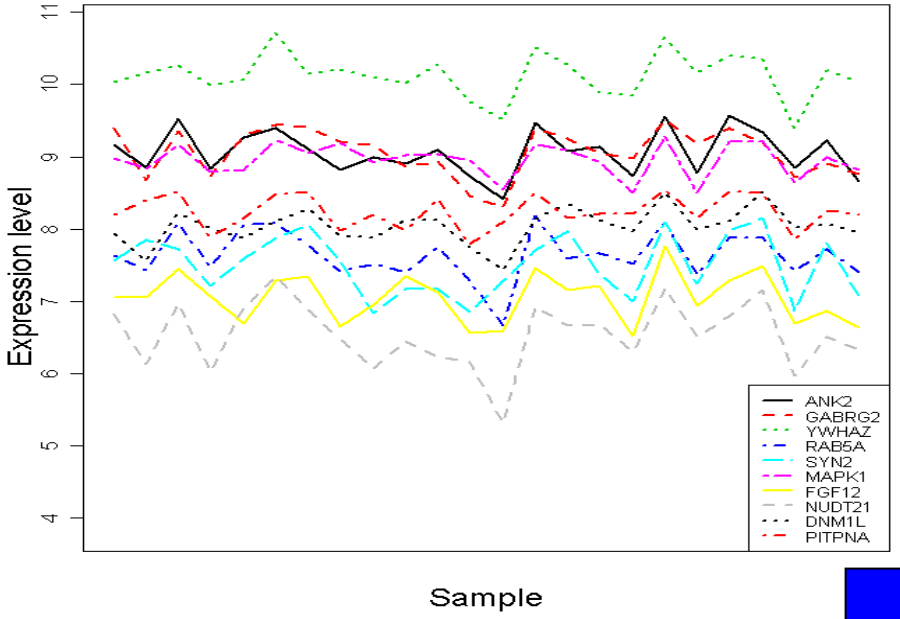
Cortex_95



Caudate nucleus



Cerebellum



Conserved modules are enriched for markers of major cell classes

Network	Module	Color	Oligodendrocytes[1]	Oligodendrocytes[2]	Astrocytes[1]	Astrocytes[2]	Neurons[1]	Neurons[2]
CTX	M9A	turquoise	138 / 352 p = 8.1e-71	47 / 60 p = 8.3e-43	NS	NS	NS	NS
CTX95	M9B	turquoise	63 / 256 p = 4.3e-37	29 / 55 p = 1.1e-27	NS	NS	NS	NS
CN	M9C	turquoise	93 / 352 p = 8.8e-66	29 / 60 p = 2.1e-28	NS	NS	NS	NS
CB	M9D	turquoise	58 / 352 p = 5.6e-38	30 / 60 p = 2.2e-35	NS	NS	NS	NS
CTX	M15A	brown	NS	NS	220 / 554 p = 2.0e-122	19 / 32 p = 5.7e-14	NS	NS
CTX95	M15B	brown	NS	11 / 55 p = 3.2e-03	145 / 430 p = 5.3e-79	13 / 27 p = 7.2e-09	NS	NS
CN	M15C	brown	NS	NS	99 / 554 p = 2.5e-61	14 / 32 p = 6.6e-14	NS	NS
CB	M15D	brown	NS	9 / 60 p = 1.0e-04	102 / 554 p = 6.1e-69	16 / 32 p = 1.1e-17	NS	NS
CTX	M16A	blue	NS	NS	NS	NS	164 / 709 p = 8.4e-30	14 / 50 p = 2.0e-03
CTX95	M16B	blue	NS	NS	NS	NS	68 / 546 p = 4.0e-15	7 / 40 p = 2.4e-02
CN	M16C	blue	NS	NS	NS	NS	64 / 709 p = 1.7e-14	11 / 50 p = 4.4e-06
CB	M16D	blue	NS	NS	NS	NS	50 / 709 p = 5.4e-05	7 / 50 p = 4.1e-02

Network	Module	Oligodendrocytes[3]	Astrocytes[4]	Synaptic proteins[5]	Synaptic proteins[6]
CTX	M9A	128/458 p=5.5e-46	NS	NS	NS
CTX_95	M9B	49/358 p=1.4e-16	4/18 p=4.6e-02	NS	NS
CN	M9C	76/458 p=2.3e-37	NS	NS	NS
CB	M9D	43/458 p=1.0e-17	NS	NS	NS
CTX	M15A	NS	15/22 p=2.6e-12	NS	NS
CTX_95	M15B	NS	12/18 p=1.9e-10	NS	NS
CN	M15C	NS	9/22 p=1.5e-08	NS	NS
CB	M15D	NS	14/22 p=1.7e-17	NS	NS
CTX	M16A	NS	NS	117/580 p=6.9e-16	23/60 p=1.6e-08
CTX_95	M16B	NS	NS	46/490 p=4.7e-06	10/55 p=1.6e-03
CN	M16C	NS	NS	55/580 p=3.1e-13	13/60 p=4.1e-07
CB	M16D	NS	NS	53/580 p=4.2e-09	15/60 p=4.0e-08

[1] Cahoy, J.D. *et al. J Neurosci* **28**: 264-78 (2008)

[2] Lein, E.S. *et al. Nature* **445**: 168-76 (2007)

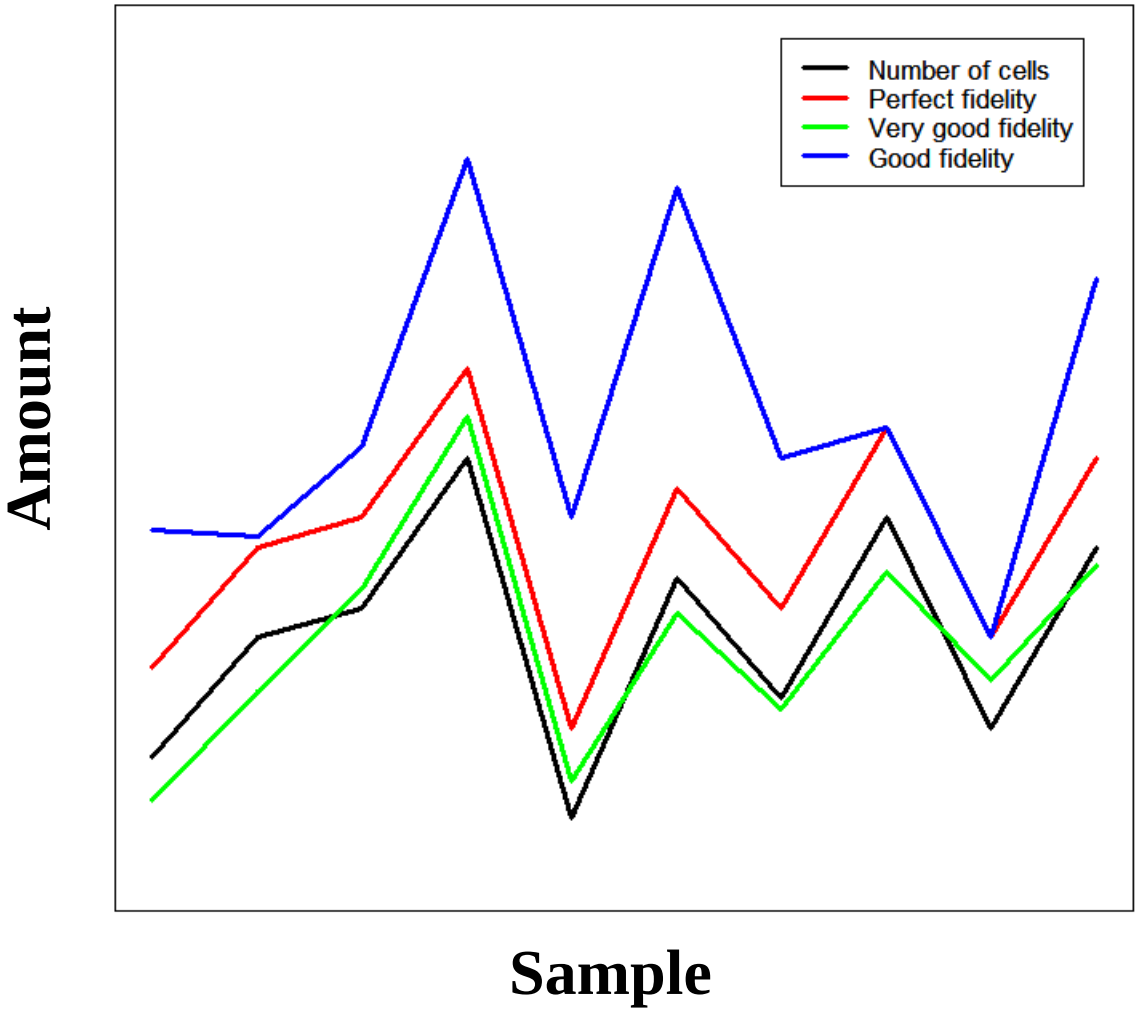
[3] Nielsen, J.A. *et al. J Neurosci* **26**: 9881-9891 (2006)

[4] Bachoo, R.M. *et al. PNAS* **101**: 8384-8389 (2004)

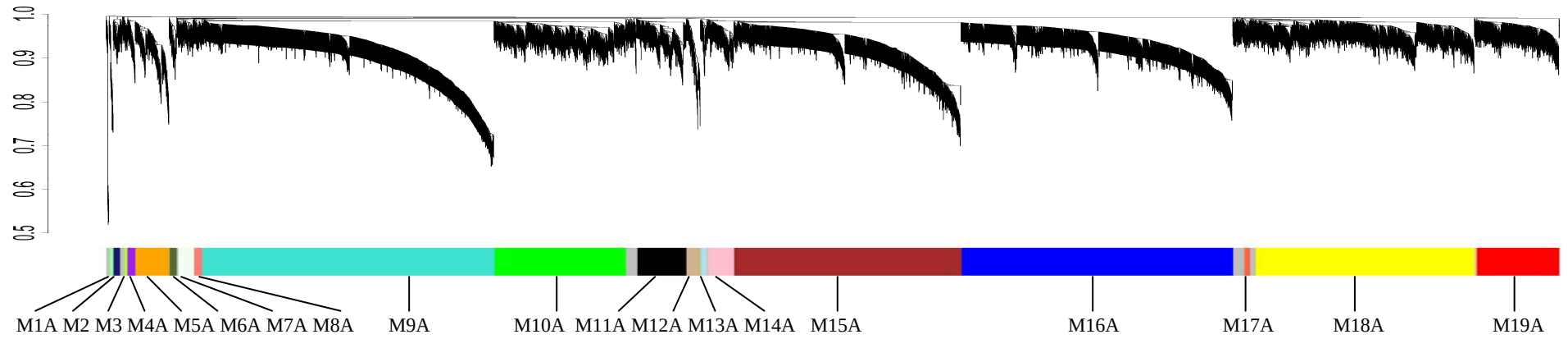
[5] Genes2Cognition Consortium

[6] Morciano, M. *et al. J Neurochem* **95**: 1732-1745 (2005)

Thought experiment

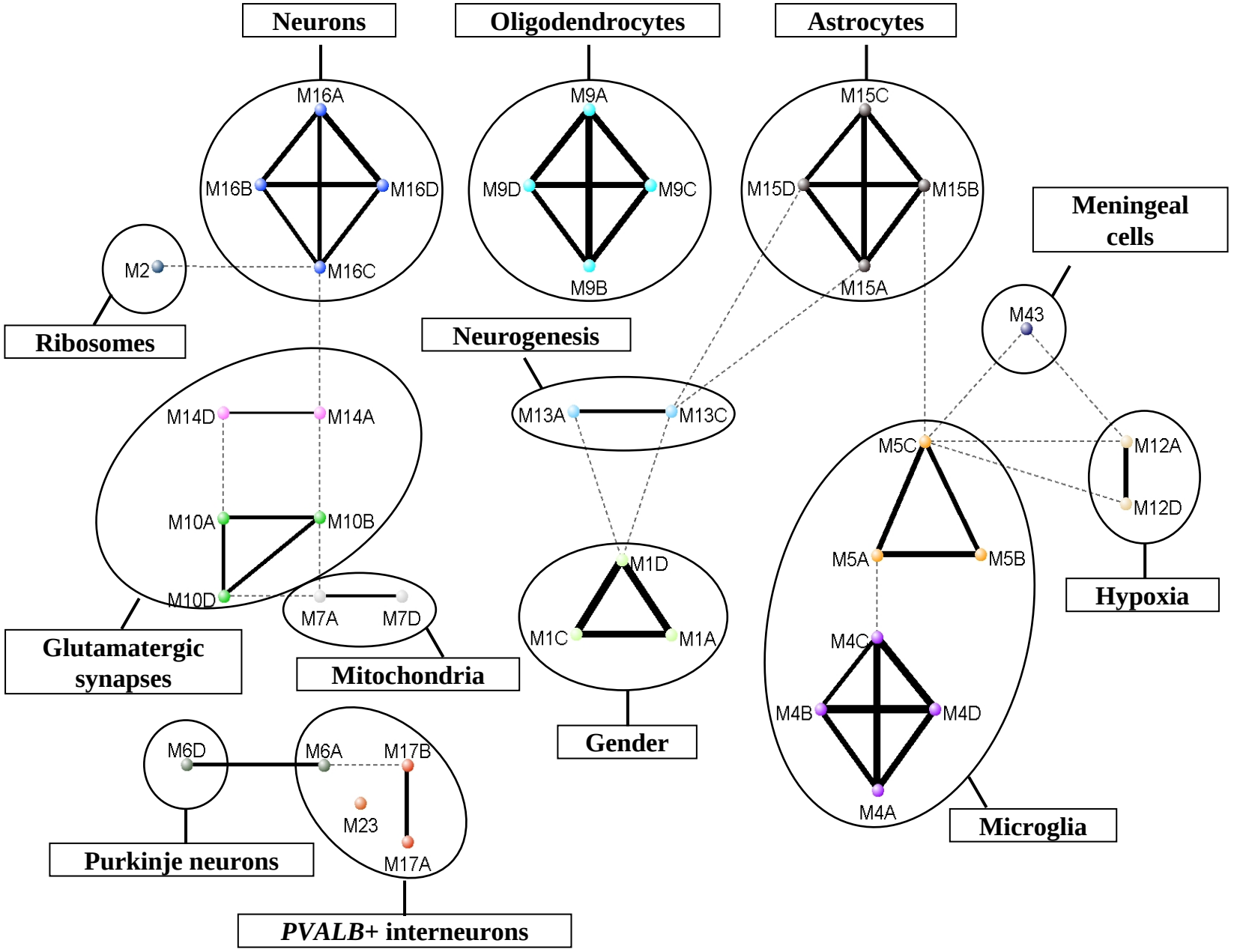


The cortical transcriptome is organized into functional modules



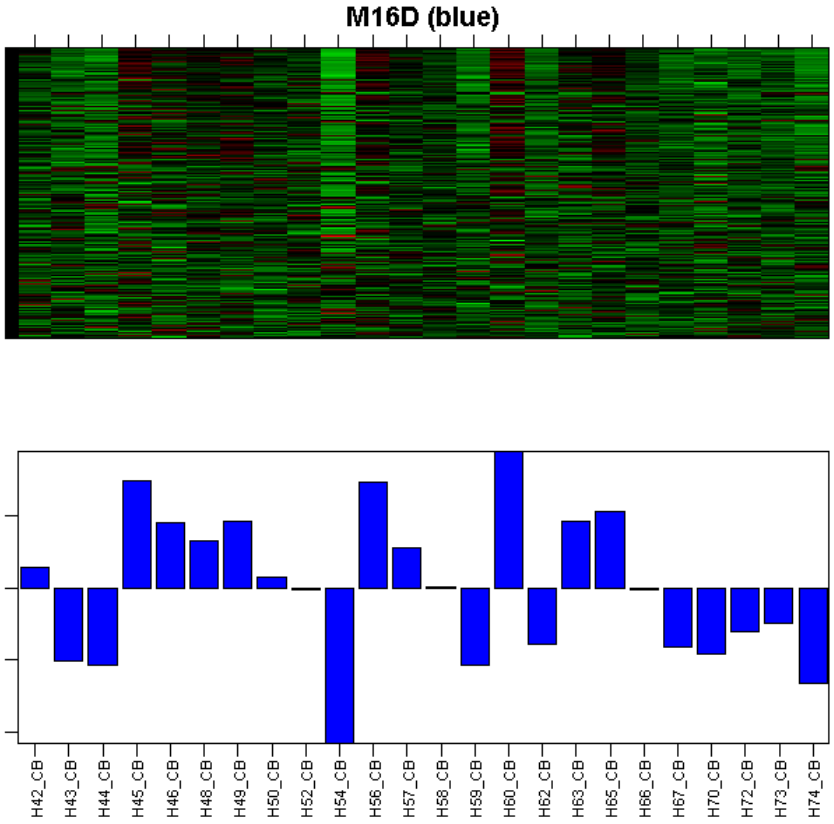
Module	Color	Interpretation	Characterization method	P-value
M1A	palegreen	Gender	Correlation with gender status	4.8e-21
M2	midnightblue	Ribosome	Gene ontology	2.3e-64
M3	greenyellow	?	N/A	N/A
M4A	purple	Immune response: microglia	Gene ontology	2.6e-56
M5A	orange	Immune response: microglia	Gene ontology	9.8e-42
M6A	darkolivegreen	PVALB+ interneuron	PVALB kme	5.3e-15
M7A	honeydew	Mitochondria	Gene ontology	2.0e-16
M8A	salmon	?	N/A	N/A
M9A	turquoise	Oligodendrocyte	Enrichment analysis with cell-type markers	8.1e-71
M10A	green	Glutamatergic synapse	Enrichment analysis with synaptic proteins	5.3e-18
M11A	black	?	N/A	N/A
M12A	tan	Hypoxia	Enrichment analysis with hypoxia genes	5.6e-13
M13A	powderblue	Neurogenesis	Gene ontology	5.0e-03
M14A	pink	Glutamatergic synapse	Enrichment analysis with synaptic proteins	1.6e-20
M15A	brown	Astrocyte	Enrichment analysis with cell-type markers	2.0e-122
M16A	blue	Neuron	Enrichment analysis with cell-type markers	8.4e-30
M17A	tomato	PVALB+ interneuron	PVALB kme	2.0e-06
M18A	yellow	Metabolism	Gene ontology	2.1e-07
M19A	red	Membrane/signal transduction	Gene ontology	6.8e-08

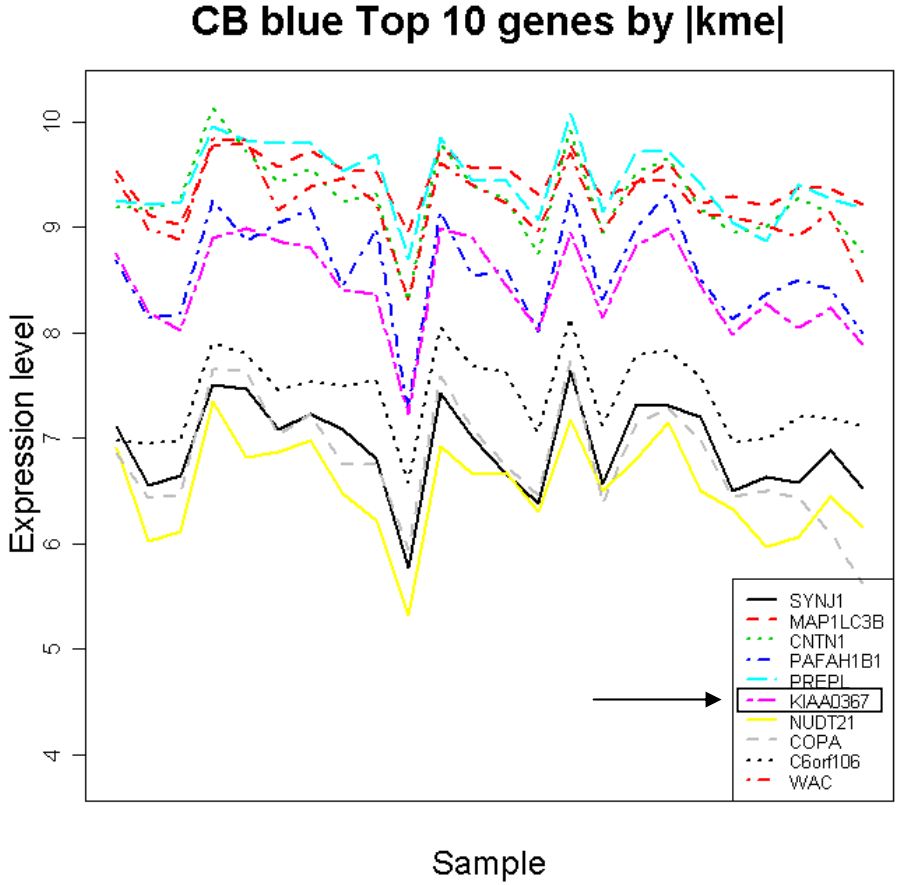
Modules are organized into a functional meta-network



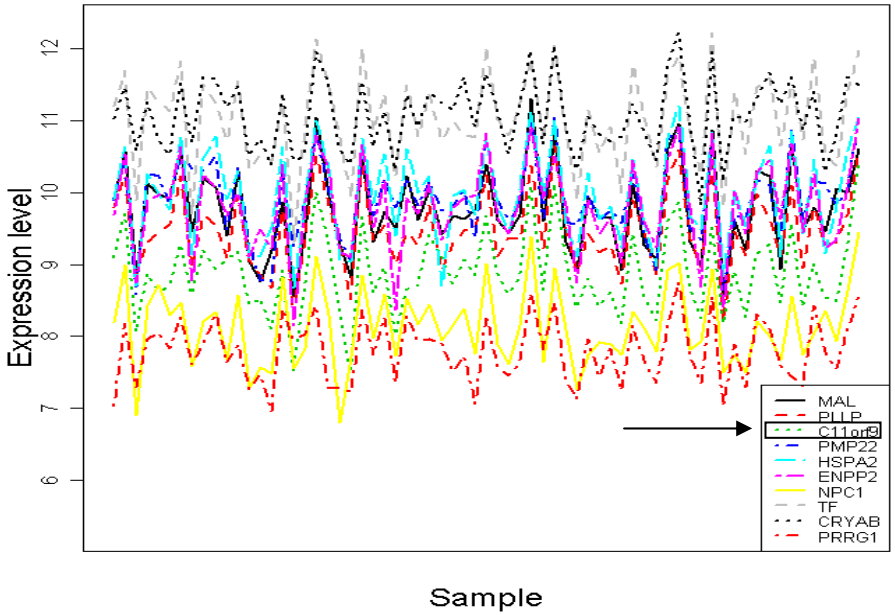
Applications

1) Context-specific annotation for genes expressed in the human brain ("guilt-by-association")
   - Rationale: genes with the strongest evidence of membership for the same module are likely to be driven by the same underlying factors


M16D (blue)
0.2
0.0
-0.2
-0.4
H42_CB
H43_CB
H44_CB
H45_CB
H46_CB
H48_CB
H49_CB
H50_CB
H52_CB
H54_CB
H56_CB
H57_CB
H58_CB
H59_CB
H60_CB
H62_CB
H63_CB
H65_CB
H66_CB
H67_CB
H70_CB
H72_CB
H73_CB
H74_CB

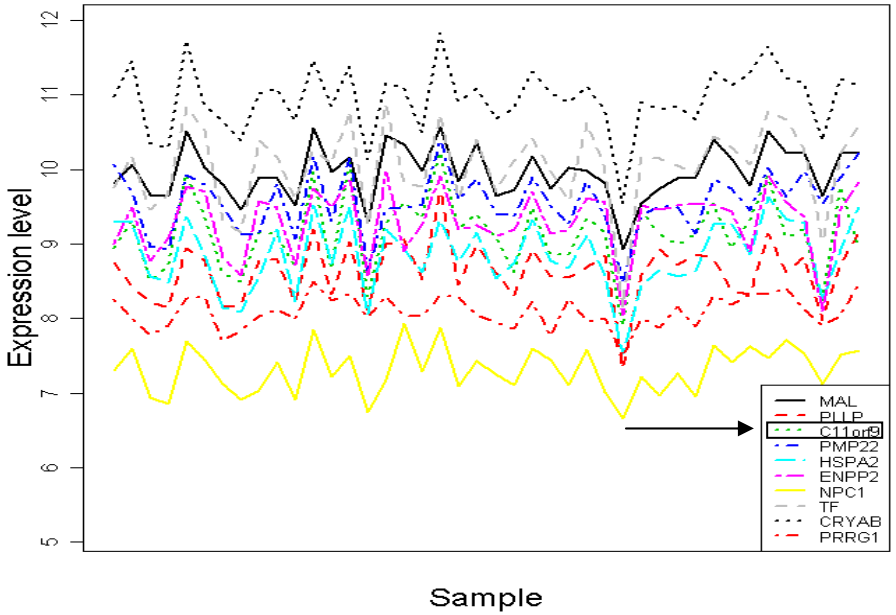

CB blue Top 10 genes by |kme|
Expression level
4
5
6
7
8
9
10
Sample
SYNJ1
MAP1LC3B
CNTN1
PAFAH1B1
PREPL
KIAA0367
NUDT21
COPA
C6orf106
WAC

Cortex



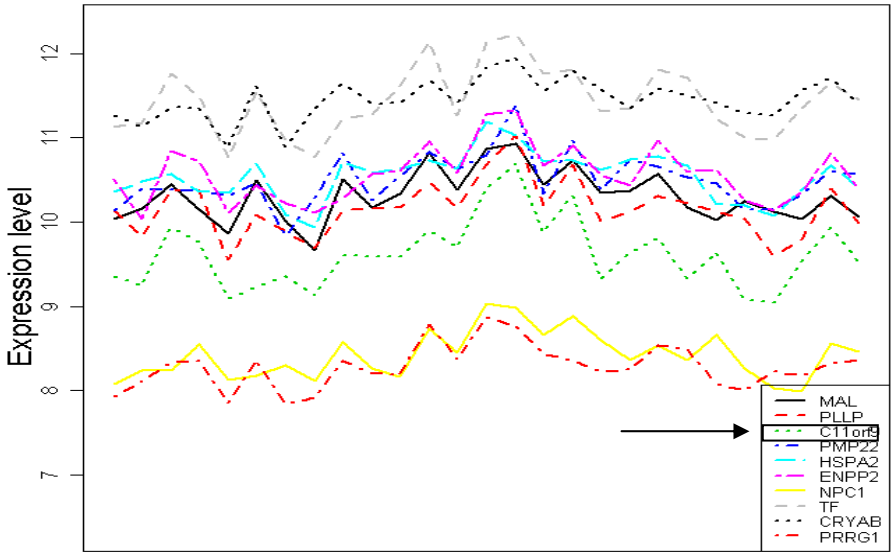
Sample

Cortex_95



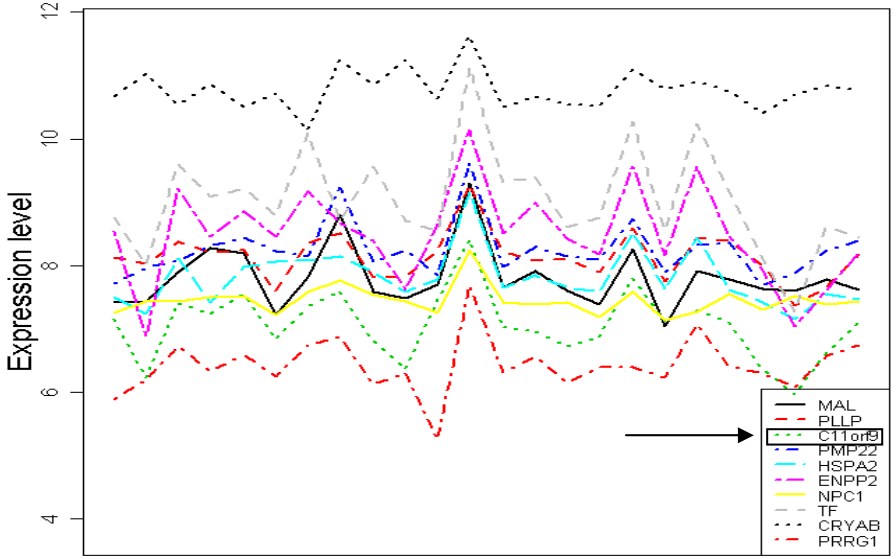
Sample

Caudate nucleus



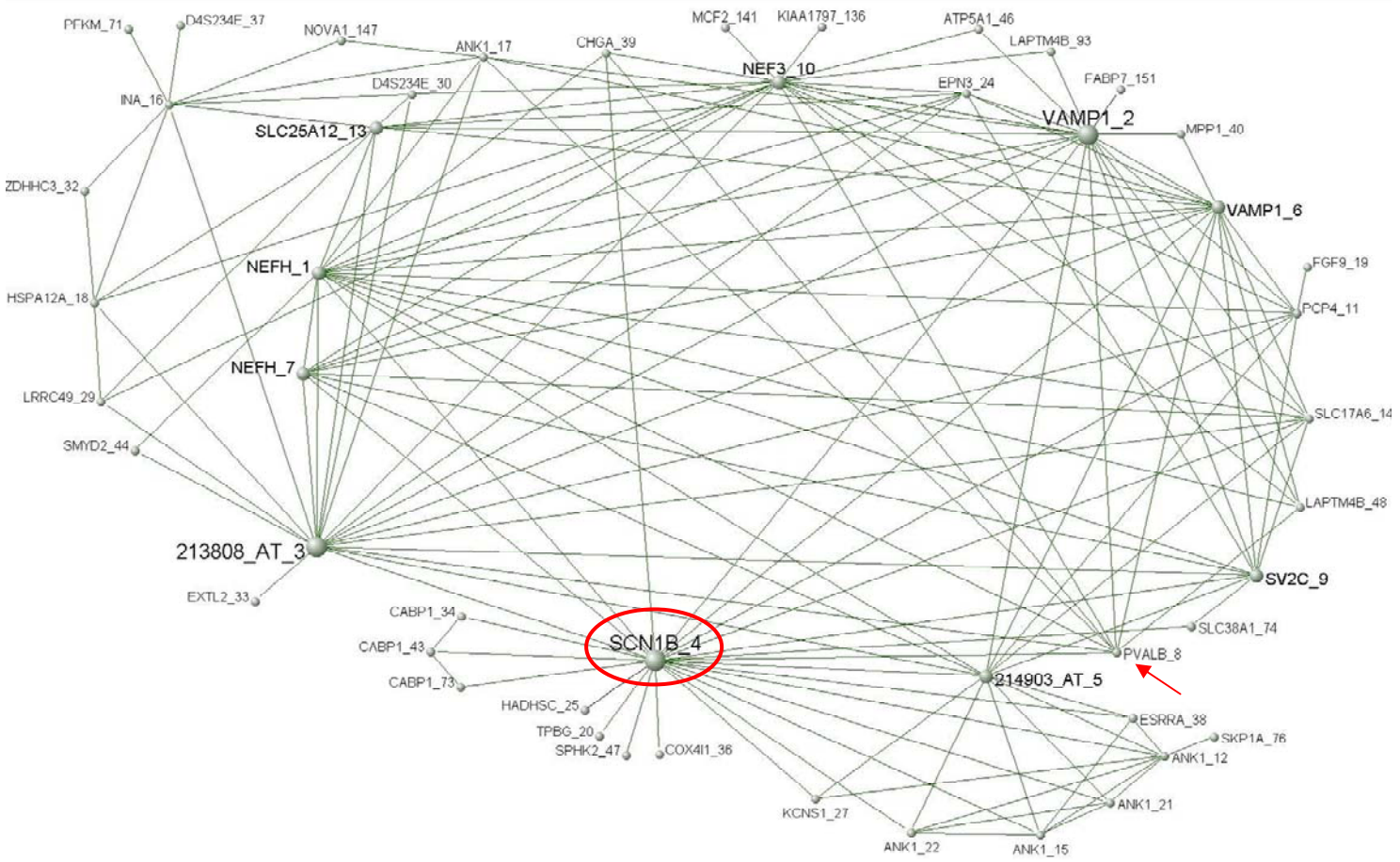
Sample

Cerebellum



Sample

M6A: *PVALB*+ interneurons



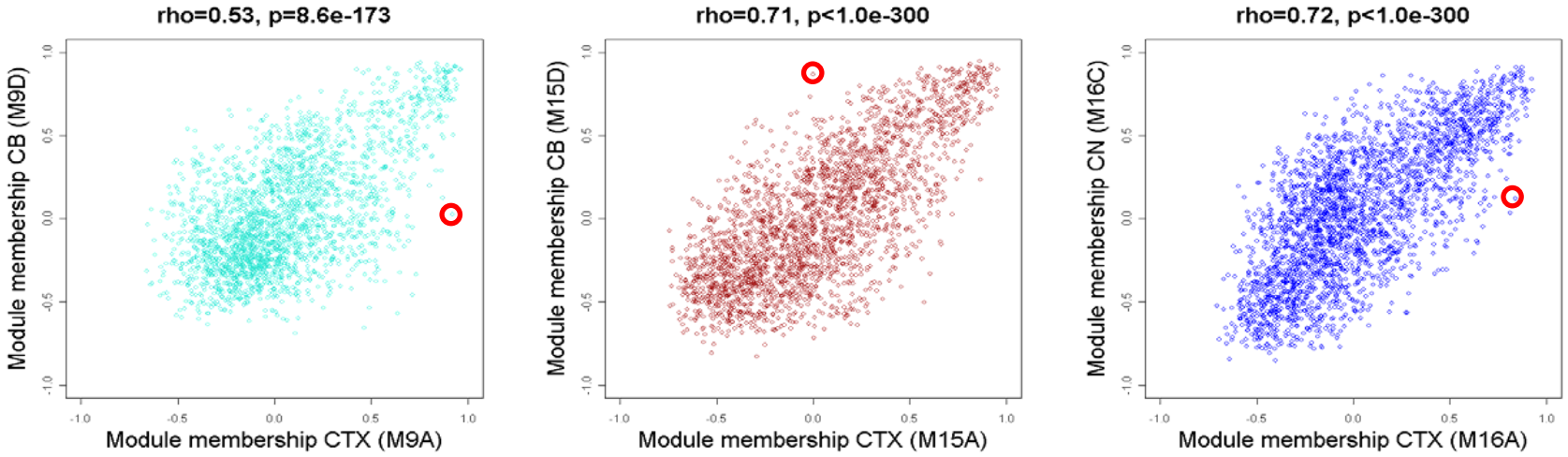
letter

Febrile seizures and generalized epilepsy associated with a mutation in the Na$^+$-channel β1 subunit gene *SCN1B*

Robyn H. Wallace[1,2], Dao W. Wang[3], Rita Singh[4,5], Ingrid E. Scheffer[4,5,6], Alfred L. George, Jr.[3], Hilary A. Phillips[1], Kathrin Saar[7], Andre Reis[7,8], Eric W. Johnson[9], Grant R. Sutherland[1,2], Samuel F. Berkovic[4,5] & John C. Mulley[1,10]

Applications

1) Context-specific annotation for genes expressed in the human brain ("guilt-by-association")
2) *In silico* comparisons of cellular specificity of gene expression across brain regions
   - Rationale: genes with the most significant differences in membership for cell-type modules between brain regions imply differences in the cellular specificity and/or consistency of gene expression

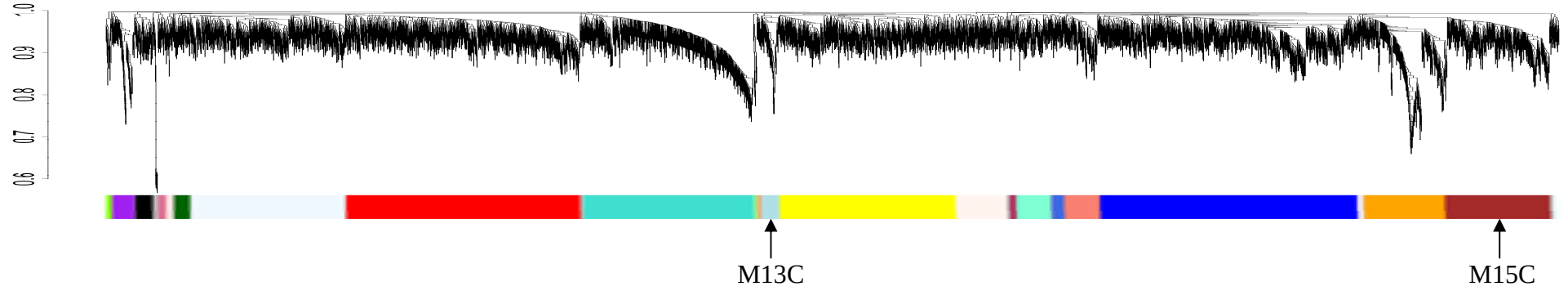


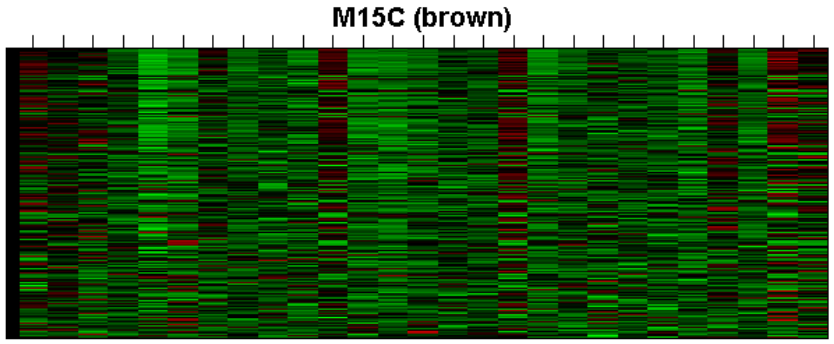
Network 1	Network 2	Probe set	Gene	Module	MM ntwk 1	MM ntwk 2	P-value	Expr ntwk 1	Expr ntwk 2	P-value
CN	CB	212754_s_at	*MON2*	M9	0.761	-0.568	3.8e-08	8.188	8.365	2.4e-02
CN	CB	208442_s_at	*ATM /// LOC651610*	M9	0.794	-0.352	1.2e-06	6.321	6.856	7.5e-04
CN	CB	206135_at	*ST18*	M9	0.877	-0.055	2.0e-06	6.895	10.033	<2.2e-16
CTX	CB	206135_at	*ST18*	M9	0.924	-0.055	2.9e-11	8.062	10.033	<2.2e-16
CTX	CB	206403_at	*ZNF536*	M9	0.922	0.029	4.1e-10	8.143	9.176	3.8e-15
CTX	CB	205425_at	*HIP1*	M9	0.834	-0.022	1.2e-06	6.427	5.789	1.7e-06
CTX	CB	217359_s_at	*NCAM1*	M9	0.870	0.126	1.6e-06	8.179	7.149	<2.2e-16
CTX	CB	205710_at	*LRP2*	M9	0.858	0.098	2.3e-06	6.105	4.562	6.5e-10
CTX	CB	212096_s_at	*MTUS1*	M9	0.835	0.021	2.5e-06	9.473	10.915	<2.2e-16
CTX	CN	218648_at	*CRTC3*	M9	0.818	-0.144	6.4e-08	9.131	9.854	<2.2e-16
CN	CB	204424_s_at	*LMO3*	M15	-0.241	0.868	1.5e-07	11.445	8.995	<2.2e-16
CN	CB	206915_at	*NKX2-2*	M15	-0.383	0.784	1.1e-06	7.301	6.901	6.4e-03
CN	CB	218656_s_at	*LHFP*	M15	0.753	-0.434	1.3e-06	7.936	9.743	<2.2e-16
CTX	CB	204424_s_at	*LMO3*	M15	0.001	0.868	1.5e-07	11.642	8.995	<2.2e-16
CTX	CB	206878_at	*DAO*	M15	-0.102	0.827	3.5e-07	6.419	8.600	<2.2e-16
CTX	CB	202834_at	*AGT*	M15	0.907	0.237	4.5e-07	9.687	10.484	7.5e-09
CTX	CB	206243_at	*TIMP4*	M15	-0.108	0.796	1.9e-06	4.194	8.867	<2.2e-16
CTX	CN	201976_s_at	*MYO10*	M15	0.856	0.085	6.1e-07	8.926	9.785	4.1e-15
CTX	CN	203786_s_at	*TPD52L1*	M15	0.756	-0.191	8.4e-07	9.741	11.340	<2.2e-16
CTX	CN	206465_at	*ACSBG1*	M15	0.924	0.449	2.1e-06	8.328	9.424	6.8e-13
CN	CB	203687_at	*CX3CL1*	M16	0.861	-0.303	7.0e-08	7.985	5.968	<2.2e-16
CN	CB	214652_at	*DRD1*	M16	0.816	-0.399	1.5e-07	8.022	3.968	<2.2e-16
CTX	CB	204338_s_at	*RGS4*	M16	0.872	-0.158	2.5e-09	10.074	5.595	<2.2e-16
CTX	CB	212093_s_at	*MTUS1*	M16	0.091	0.897	5.7e-08	8.173	9.614	<2.2e-16
CTX	CB	214564_s_at	*PCDHGC3*	M16	-0.150	0.819	2.1e-07	6.626	7.510	1.1e-16
CTX	CB	208485_x_at	*CFLAR*	M16	-0.141	0.821	2.3e-07	6.550	8.148	<2.2e-16
CTX	CN	202543_s_at	*GMFB*	M16	0.894	0.162	9.4e-08	7.547	6.735	2.0e-11
CTX	CN	210735_s_at	*CA12*	M16	-0.027	0.842	1.6e-07	6.009	8.670	<2.2e-16
CTX	CN	201831_s_at	*PAK1 /// VDP*	M16	0.836	0.029	8.4e-07	7.454	6.703	4.0e-10
CTX	CN	221487_s_at	*ENSA*	M16	0.839	0.053	1.2e-06	8.155	6.796	<2.2e-16

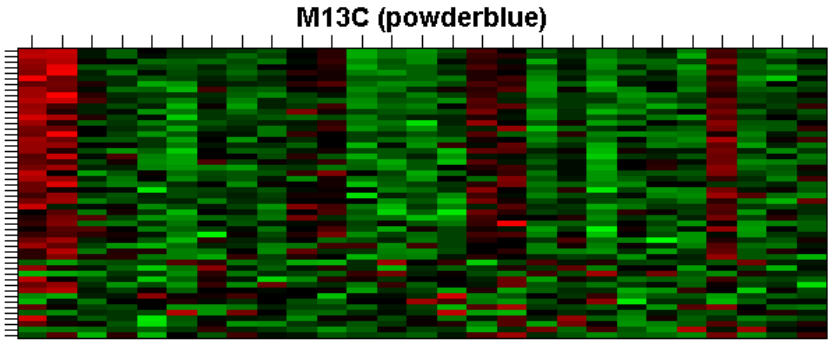
Applications

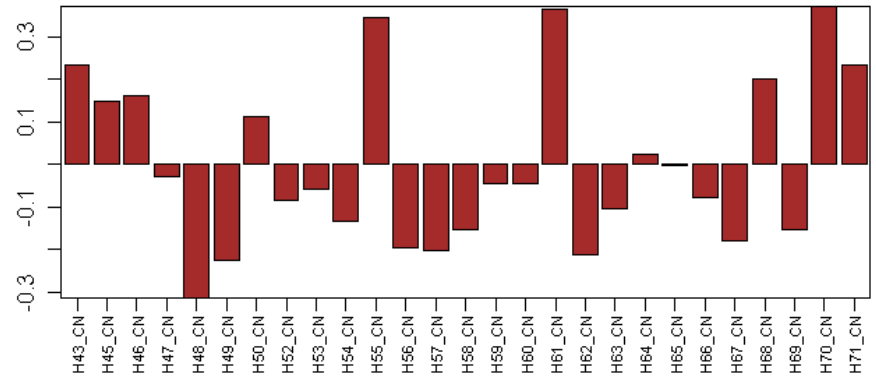
1) Context-specific annotation for genes expressed in the human brain ("guilt-by-association")
2) *In silico* comparisons of cellular specificity of gene expression across brain regions
3) Cellular phenotype discovery
   - Rationale: unsupervised analysis of gene coexpression patterns can identify novel distinctions among cell types within brain regions

Caudate nucleus

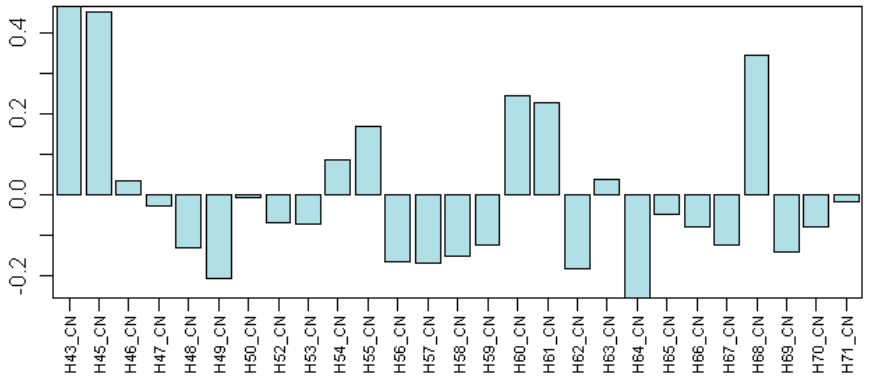








Astrocytes[1]	Astrocytes[2]	Astrocytes[3]
99 / 554 (p=2.5e-61)	14 / 32 (p=6.6e-14)	9 / 22 (p=1.5e-08)



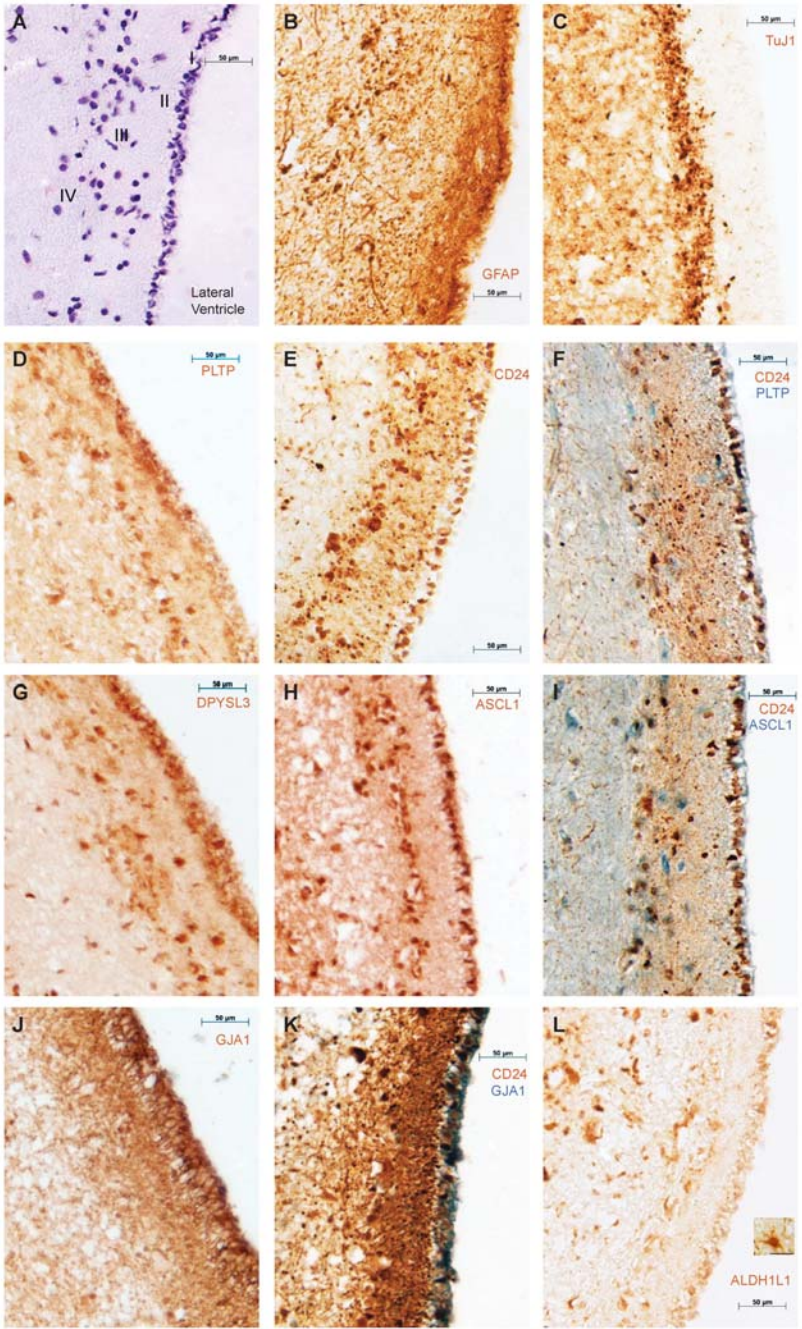
Astrocytes[1]	Astrocytes[2]	Astrocytes[3]
45 / 554 (p=1.9e-24)	8 / 32 (p=6.6e-08)	7 / 22 (p=1.0e-07)

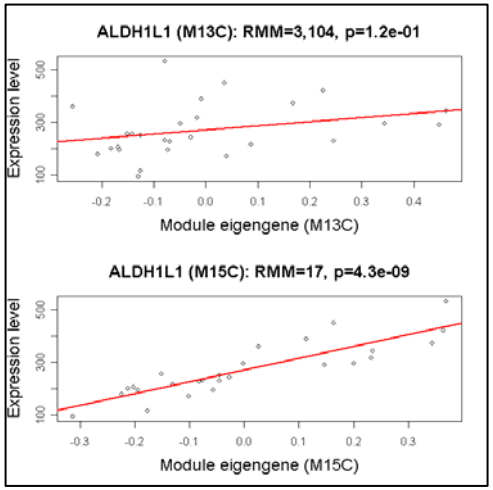
[1] Cahoy, J.D. *et al. J Neurosci* **28**: 264-78 (2008)

[2] Lein, E.S. *et al. Nature* **445**: 168-76 (2007)

[3] Bachoo, R.M. *et al. PNAS* **101**: 8384-8389 (2004)

M13C (CN) identifies genes that are coexpressed in the SVZ

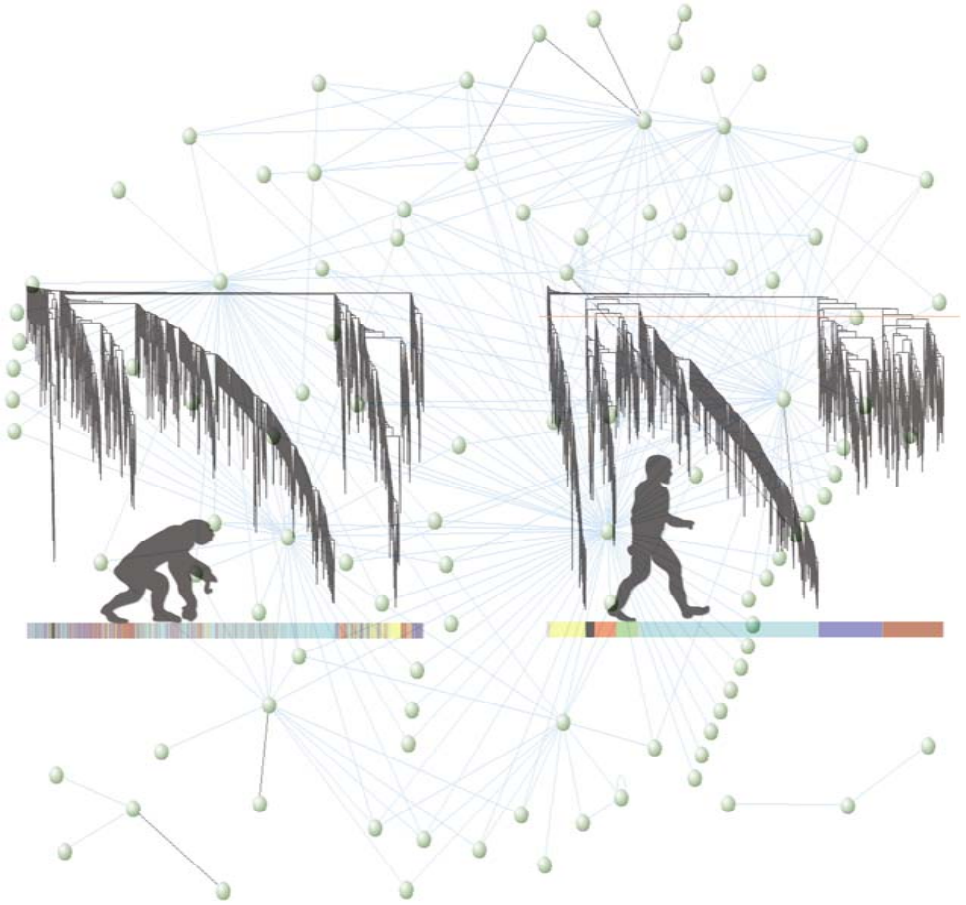




Conclusions

- The human brain transcriptome is organized into modules of coexpressed genes
  - Many modules are reproducible across microarrays, individuals, and brain regions
- Several highly conserved modules are enriched for markers of major cell classes
  - 'Core' transcriptional programs for neurons, oligodendrocytes, astrocytes, and microglia
  - Context-specific annotation for thousands of genes expressed in the human brain
- Potential to leverage consistency of k_{me} for comparisons with other conditions of interest (e.g. disease)

Acknowledgements



Dan Geschwind
Steve Horvath
Gena Konopka
Peter Langfelder

Collaborators:
Dr. Kazuya Iwamoto
Dr. Tadafumi Kato

The Geschwind lab